# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| January 2014 | Journal Article | January 2014- August 2014 |

**4. TITLE AND SUBTITLE**

Modeling RP-1 Fuel Advanced Distillation Data using Comprehensive Two-Dimensional Gas Chromatography coupled with Time-of-Flight Mass Spectrometry and Partial Least Squares Analysis

**5a. CONTRACT NUMBER**
In-House

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Kehimkar, B., B. Parsons, J. Hoggard, M. Billingsley, T. Bruno, R. Synovec

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**
Q0A4

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Air Force Research Laboratory (AFMC)
AFRL/RQRC
10 E. Saturn Blvd.
Edwards AFB, CA, 93524-7680

**8. PERFORMING ORGANIZATION REPORT NO.**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Research Laboratory (AFMC)
AFRL/RQR
5 Pollux Drive.
Edwards AFB, CA, 93524-7048

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
**AFRL-RQ-ED-JA-2014-175**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**13. SUPPLEMENTARY NOTES**
Journal article published in the Analytical and Bioanalytical Chemistry, Vol. #407, Issue #1 Jan 2015.
PA Case Number: #14342; Clearance Date: 07 July 14.


## 14. ABSTRACT

Recent efforts in predicting rocket propulsion (RP-1) fuel performance through modeling put greater emphasis on obtaining detailed and accurate fuel properties, as well as to elucidate the relationships between fuel composition and their properties. Herein, we study multidimensional chromatographic data obtained utilizing the instrumental platform that included comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC × GC –TOFMS) to analyze RP-1 fuels. For GC × GC separations, RTX-wax (polar stationary phase) and RTX-1 (non-polar stationary phase) columns were implemented for the primary and secondary dimensions, respectively, to separate the chemical compound classes (alkanes, cycloalkanes, aromatics, etc), providing a significant level of chemical compositional information. The GC × GC – TOFMS data were analyzed using partial least-squares regression (PLS) chemometric analysis, specifically to model and predict advanced distillation curve (ADC) data for ten RP-1 fuels that were previously analyzed using the ADC method. The PLS modeling provides insight into the chemical species that impact the observed changes in the previously collected ADC data. The PLS modeling correlates compositional information found in the GC × GC – TOFMS chromatograms of each RP-1 fuel, and their respective ADC, and allows prediction of the ADC for each RP-1 fuel with good precision and accuracy. The predictive power of the overall method via PLS modeling was assessed using leave-one-out cross-validation (LOOCV) yielding root-mean-square error of cross-validation (RMSECV) with low values, typically below 2.0 °C, at each % distilled measurement point during the ADC analysis.

**15. SUBJECT TERMS**

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| | | | | | Matt Billingsley |
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | **19b. TELEPHONE NO** *(include area code)* |
| **Unclassified** | **Unclassified** | **Unclassified** | SAR | 34 | 661-275-5885 |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. 239.18**

# Modeling RP-1 Fuel Advanced Distillation Data using Comprehensive Two-Dimensional Gas Chromatography coupled with Time-of-Flight Mass Spectrometry and Partial Least Squares Analysis

Benjamin Kehimkar,[a] Brendon A. Parsons,[a] Jamin C. Hoggard,[a] Matthew C. Billingsley,[b] Thomas J. Bruno,[c] and Robert E. Synovec[a,*]

[a] Department of Chemistry, Box 351700, University of Washington
Seattle, WA 98195, USA

[b] Air Force Research Laboratory/RQRC, 10 E Saturn Blvd.
Edwards AFB, CA 93524, USA

[c] Applied Chemicals and Materials Division, National Institute of Standards and Technology
Boulder, CO 80305, USA

* Corresponding author at: Department of Chemistry, University of Washington, Seattle, WA 98195-1700, USA. Tel.:+1 206 685 2328; fax: +1 206 685 8665.
E-mail address: synovec@chem.washington.edu (R. E. Synovec).

Prepared for submission to **Analytical and Bioanalytical Chemistry**

**Special Issue on 'Multidimensional Chromatography'**

May 7, 2014

## Abstract

Recent efforts in predicting rocket propulsion (RP-1) fuel performance through modeling put greater emphasis on obtaining detailed and accurate fuel properties, as well as to elucidate the relationships between fuel composition and their properties. Herein, we study multidimensional chromatographic data obtained utilizing the instrumental platform that included comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC × GC –TOFMS) to analyze RP-1 fuels. For GC × GC separations, RTX-wax (polar stationary phase) and RTX-1 (non-polar stationary phase) columns were implemented for the primary and secondary dimensions, respectively, to separate the chemical compound classes (alkanes, cycloalkanes, aromatics, etc), providing a significant level of chemical compositional information. The GC × GC – TOFMS data were analyzed using partial least-squares regression (PLS) chemometric analysis, specifically to model and predict advanced distillation curve (ADC) data for ten RP-1 fuels that were previously analyzed using the ADC method. The PLS modeling provides insight into the chemical species that impact the observed changes in the previously collected ADC data. The PLS modeling correlates compositional information found in the GC × GC – TOFMS chromatograms of each RP-1 fuel, and their respective ADC, and allows prediction of the ADC for each RP-1 fuel with good precision and accuracy. The predictive power of the overall method via PLS modeling was assessed using leave-one-out cross-validation (LOOCV) yielding root-mean-square error of cross-validation (RMSECV) with low values, typically below 2.0 °C, at each % distilled measurement point during the ADC analysis.

**Keywords**: GC × GC – TOFMS, partial least squares (PLS) analysis, advanced distillation curve (ADC), two-dimensional, gas chromatography, RP-1 fuel.

## Introduction

The chemical composition of a kerosene fuel, though complex, holds a key to understanding and altering the physical properties and performance of the fuel [1–7]. Achieving fine control over the chemical composition can be a difficult task. It has become increasingly important to achieve further insight into fuel composition, as well as the sources of variation in the fuel composition to both maintain and control fuel performance, as well as to assess the performance of "field" fuels [1–5]. Fuel performance is inextricably tied to characterization, and the advanced distillation curve (ADC) method has demonstrated itself as a well suited approach for the analysis and characterization of complex fuels [8–10]. The ADC method is a state-of-the-art approach to very accurately and precisely analyze the boiling curve of complex liquids. Samples (i.e., distillation fractions) may be obtained during the distillation, and can be further analyzed both qualitatively and quantitatively.

The ADC method was pioneered by Bruno and co-workers [2, 5, 8–20]. Briefly, the apparatus for the ADC method utilizes a round-bottom flask connected to an air cooled condenser, a receiver adapter and a calibrated volume receiver. The flask is encased with a heater in an aluminum jacket. Inside the flask are two thermocouples, suspended using a centering adapter: one thermocouple measures the temperature of the liquid analyzed, and the other thermocouple measures the temperature in the headspace above the liquid being distilled. Three bore scope ports are strategically located to inspect both the liquid and the thermocouples inside the apparatus. The flask is connected to an air cooled condenser chilled with a vortex tube, wherein the distillate condenses. The condenser, in turn, is connected to a special adapter where the drops of distillate fall into a small 0.05 mL "hammock." With the use of a syringe, the distillate may be sampled from the hammock for further analysis including, but not limited to,

gas chromatography (GC) [9–15], infrared spectroscopy [12], and measurements of enthalpy of combustion [14]. After the adapter, the distillate reaches the calibrated volume receiver. More recently, a variation of the ADC method apparatus was implemented that controls the internal pressure, preventing sample degradation due to reactions that may potentially occur at high temperatures when analyzing samples containing low-volatility compounds [17]. This feature was achieved by sealing every connection between parts of the apparatus and using a commercial pressure controller. Sampling is performed with a reduced pressure balance syringe.

The ADC method has been instrumental in the study of a variety of complex liquid samples including, but not limited to, crude oil [12], gasoline [16], biodiesel fuel [17, 19], jet fuel [5, 10, 11], motor oil [18], and rocket propellant (RP) [2, 9, 10, 13–15, 20]. The ADC method can be used not only to provide information regarding sample composition, but also to study the thermodynamic and physical properties, chemical properties such as corrosive effects [12], enthalpy of combustion (through the use of each distillate fraction to determine the overall enthalpy of combustion) [5, 10, 11, 15–16], and the influence of thermal stress on fuels [15]. Furthermore, the variability in fuel composition and its impact on thermophysical properties have also been investigated [20].

In conjunction with implementing the ADC method, it has become apparent that additional chemical composition information should be evaluated to strengthen and ultimately apply the information gained from ADC data. For this purpose, in this report we applied the powerful chemical analysis tool known as comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC × GC – TOFMS), using a reverse column GC × GC configuration (i.e., polar primary dimension column coupled with a non-polar secondary dimension column) [21] building from our previous study [22], to improve the

97 separation of the various compound classes (eg. alkanes, cycloalkanes, aromatics, etc), and to
98 facilitate extraction of chemical information from a set of ten RP-1 fuel samples. Using
99 chemometrics, we then explored the connection between chemical composition via GC × GC –
100 TOFMS chromatographic data and the ADC data from the RP-1 fuels. Indeed, GC × GC –
101 TOFMS is ideally suited for use in fuels analysis [21–31].
102 To help glean useful information, multivariate "chemometric" data analysis methods
103 have been developed. Chemometrics have been shown to be able to take advantage of the
104 three-way data provided by the GC × GC – TOFMS instrumental platform, to help reveal
105 similarities and/or differences between chromatograms [22–25, 32]. Partial least-squares (PLS)
106 analysis can be used to associate variance in fuel composition to measured physical properties
107 [22]. Detailed information on the theory of PLS can be found elsewhere [34–36]. In this study,
108 GC × GC – TOFMS chromatographic data of RP-1 fuels and their respective ADC data are
109 analyzed using PLS to provide useful information on chemical compounds that significantly
110 influence the RP-1 fuel properties via inspection of the linear regression vector (LRV) of each
111 PLS model. This analysis is accomplished by selecting an appropriate number of latent variables
112 (LVs) that are used to calculate loadings that capture the variance (i.e. chemical information) in
113 the GC × GC – TOFMS chromatograms that have the maximum covariance with corresponding
114 information in the ADC data set. Our goals are to demonstrate and validate the use of PLS
115 modeling, and to relate chemical information obtained from the GC × GC – TOFMS
116 chromatograms to the corresponding ADC for each RP-1 fuel, and ultimately to predict the ADC
117 temperatures of a given RP-1 fuel, without directly making those measurements [2]. This
118 chemical analysis approach has the ability to provide insight into the chemical composition
119 changes as a function of % distilled (and distillation temperature during the ADC experiment).

Eventually, this chemical analysis approach will provide insight, and to aid, in the process of optimization of fuel performance.

## Experimental

### GC × GC – TOFMS data collection

The full details on the GC × GC – TOFMS instrumental platform and methodology can be found in our previous report [22]. Ten RP-1 fuel samples were obtained from the Air Force Research Laboratory (AFRL), Edwards AFB, CA, and are listed in Table. 1. The ADC data were obtained from an earlier report [2]. The GC × GC – TOFMS instrument used was an Agilent 6890A GC with a 7683B auto-injector (Agilent Technologies, Palo Alto, CA, USA) coupled to a LECO Pegasus-III TOFMS (LECO, St. Joseph, MI, USA). Isobaric mode was used with an inlet pressure of 35 psig (241 kPa). The auto-injector was set to 1 μL injection, a 200:1 split injection with helium carrier gas was used, and acetone was used as the solvent rinse. The first GC × GC separation dimension (primary column) used a RTX-wax (polar) stationary phase, of 30 m in length, 250 μm i.d., and a 0.5 μm film. The modulation period was set to 2.5 s. The second separation dimension (secondary column) used a 1.2 m RTX-1, of 100 μm i.d., and a 0.18 μm film. The GC oven was initially set to 40 °C for 2 min and ramped to 225 °C at a rate of 6 °C/min; the final temperature was maintained for 3 min. The GC inlet was set to 225 °C and the transfer line temperature was 235 °C. The thermal modulator offset was 20 °C, with a hot pulse time of 0.59 s and a 0.35 s cool time. The secondary column oven temperature control was not used while still achieving a suitable GC × GC separation, and the secondary oven (housed in the primary oven) was left open and set at the same nominal temperature as the primary oven. The TOFMS data acquisition parameters were set with a 120 s acquisition delay, a mass channel ($m/z$) scan range of 35-334 amu, with a 100 Hz acquisition rate.

**Data analysis: PLS of GC × GC – TOFMS data**

The computer used for analysis was an Intel Core i-3-2120 @3.3 GHz, with 16.0 GB of RAM, and included a 60 GB SSD drive used for the purpose of a page disc ("fast" virtual RAM). Two replicate sets of RP-1 GC × GC – TOFMS chromatograms were analyzed as separate sets of PLS models as described below, and the results for both replicates are provided herein, overlaid in figures, similar to previous reports [22, 30]. Chromatographic runs were imported to MATLAB2009b (MathWorks, Natick MA) using the 'peg2mat' function [22, 37–38].

The GC × GC – TOFMS data underwent baseline correction using in-house software as reported previously [22], and to help save memory and computation time, the data also underwent a condensing procedure [22, 39] that included the following operations. First, the chromatographic data were binned (for 2 points in each chromatographic dimension, resulting in GC × GC – TOFMS chromatograms that are 25% of their original size). The binning also addressed any minor run-to-run misalignment in the data [39]. Second, in the TOFMS domain, omitting *m*/*z* channels that were unselective and *m*/*z* channels that do not exhibit signal greater than five times the standard deviation of baseline corrected noise (these *m*/*z* are: 35-37,43-47, 51, 58-62, 73-76, 87-90, 101-103, 115-118, 133, 207, 214-334). Third, the signal for uninformative temporal regions was set to 0, specifically, GC × GC regions dominated by column bleed or with no analyte compound signal (these regions were initially inspected while taking chromatogram variability into consideration to prevent the chance of removing compositional variation). The chromatographic and mass spectral dimensions of the GC × GC – TOFMS data for each RP-1 fuel was vectorized (from 10 fuels ×125 secondary column data points ×405 primary column data points ×148 mass channels to 10 fuels × 7,492,500 unfolded data points) prior to PLS analysis along with the ADC (in vector form) for each RP-1 fuel. PLS analysis was performed

using PLS Toolbox 6.7 (Eigenvector Research Inc., Wenatchee WA), with mean centering of the GC × GC – TOFMS data and auto scaling (subtracting the mean and dividing by the standard deviation) for the ADC temperature values.

**Data analysis: PLS of ADC data**

Using the ADC method for a RP-1 fuel analysis, the temperature is recorded at the moment a specific percentage of the fuel has been distilled (% distilled point) [2]. For this study, temperatures for the ADC method were measured at nineteen % distilled points: 0.025, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, and 90 [2]. Rather than construct a single PLS model for the entire ADC data set (simultaneously on all nineteen measurement points along the % distilled axis of the ADC for all fuels in the sample set), a series of 19 PLS models (a PLS model at each % distilled point) were produced. Performing the PLS analysis using a series of 19 models offered several key advantages. First, this approach lessened the restrictions on PLS when constructing the model(s). Second, this approach offered the ability to change the number of LVs at different % distilled points in the ADC (if necessary). Different numbers of LVs can be expected because the composition of a fuel is known to change over the course of the distillation, i.e. the GC × GC – TOFMS chromatographic data represents the initial chemical composition of a given fuel, however the composition at a given % distilled is a subset of this composition, with possibly different relative concentrations for the various compounds present. A third important advantage for constructing a series of 19 PLS models was to save computation time. Consider modeling the entire ADC data set (10 fuels × 19 % distilled points) coupled with the unfolded GC × GC – TOFMS chromatograms (as stated previously, 7,492,500 unfolded data points per fuel): PLS would require a considerable amount of computer memory (about 13 GB), and the computation time would be prohibitively long, and on some computer systems this

computational exercise would fail due to memory constraints. In contrast, applying PLS on the unfolded GC × GC – TOFMS chromatograms at one % distilled point at a time required fewer LVs and significantly less memory (around 6.5GB), and required less than a minute to compute per PLS model.

**PLS modeling of GC × GC – TOFMS and ADC data**

The PLS modeling was validated using leave-one-out-cross-validation (LOOCV). Briefly, LOOCV involves a series of PLS models from (*n*-1) samples from the original *n* sample data set, using the $n^{th}$ sample to predict values from the constructed (*n*-1) model. After all combinations are analyzed the root-mean-square of error of cross-validation of the residuals of the PLS models (RMSECV) was calculated [40]:

$$RMSECV = \sqrt{\frac{\sum(y_{cv} - y_{meas})^2}{N}} \quad (1)$$

Moreover, RMSECV results were also used to help determine the most appropriate number of LVs to use for the PLS models.

At each step in the analytical procedure, the LRVs of the PLS models were inspected to qualitatively verify that the connections the PLS models made between the chromatographic information (GC × GC – TOFMS data) and physical measurements (ADC data) were both logical, and that the LRVs from consecutive models appear continuous. Using information provided by the LRVs, identification of compounds of interest in the GC × GC – TOFMS data was performed via ChromaTOF V.3.32 (LECO Corporation, St. Joseph, MI, USA), and in-house software for nontarget PARAFAC for well resolved and unresolved peaks, respectively [26]. The NIST11 V2.0g mass spectral library (National Institute of Standards and Technology, Boulder CO, USA) was used for mass spectral identification.

## Results and discussion

A representative GC × GC – TOFMS chromatogram of an RP-1 fuel is provided in Fig.1a. In this figure the total ion current (TIC) signal is plotted for the GC × GC separation of the RP-1 fuel LB073009-08. To further demonstrate the separation power for complex samples such as RP-1, in Fig. 1b, c and d, are provided specific regions of the GC × GC separation with a representative alkane, cycloalkane, and aromatic compound indicated, respectively. Each of the representative compounds indicated also are key compounds identified in the PLS modeling that will be presented herein. In Fig. 1b is provided a region of Fig. 1a at the selective mass channel *m*/*z* 57; the highlighted peak (located at 8.75 min and 1.94 s on the primary and secondary dimensions, respectively) has been identified as decane. In Fig. 1c is provided a region of Fig. 1a at the selective mass channel *m*/*z* 136; the highlighted peak (located at 15.00 min and 1.17 s on the primary and secondary dimensions, respectively) has been identified as the adamantane. Finally, in Fig. 1d is provided a region of Fig. 1a at the selective mass channel *m*/*z* 105; the highlighted peak (located at 19.29 min and 0.95 s on the primary and secondary dimensions, respectively) has been identified as methylbutylbenzene.

The previously measured ADC data for all ten RP-1 fuels are provided in Fig. 2a [2]. The measured ADC data were obtained at a % distilled range from 0.025% to 90%. The recorded temperatures for the ADC data set range from 207.2°C to 213.5°C at 0.025% distilled, to 235.9°C to 258.1°C at 90% distilled. At various % distilled values the ADC for several fuel pairs cross one another, which may potentially make the PLS modeling of ADC data more challenging. For clarity, in Fig. 2b two representative ADCs are provided that approximately span the range of temperatures at each % distilled. In Fig. 2b, RP-1 fuel LB073009-06

represents the highest measured temperatures for the ten fuels, while RP-1 fuel LB073009-02 exhibited some of the lowest recorded temperatures.

For comparison to Fig 2b, the ADCs for RP-1 fuels LB073009-06 and LB073009-02 predicted using PLS during the LOOCV procedure are provided in Fig. 2c. Figs. 2b and 2c are qualitatively very similar indicating the ability of the PLS models to accurately predict fuel physical data, but in order to obtain a more quantitative evaluation of the accuracy of the PLS modeling, residuals for each ADC were calculated at each % distilled value. The ADC residuals were obtained by subtracting a measured ADC from the ADC predicted using PLS. The residuals imply an accuracy of the PLS modeling to within +/- 2.5 °C range, which is deemed reasonable for this initial study.

Examination of the LRVs of the PLS models provide additional information, complementary to the ADCs predicted from the PLS models. In Fig. 3a-c, three of the nineteen LRVs are provided (one for each PLS model constructed, other LRVs omitted for brevity): one LRV from the beginning (0.025% distilled), middle (45% distilled), and end (90% distilled) of the ADC. Through inspection of the positive LRV values, the corresponding peaks tend to be analyte compounds eluting after ~10 min for alkanes, after ~15 min for the cycloalkanes, and di- and tri-cycloalkanes, and after ~17 min for aromatic groups to a lesser extent. These results in the LRVs display a general pattern that the less volatile compounds contribute positively to an ADC, suggesting less volatile compounds increase the overall predicted temperature of the ADC at a given % distilled point. As the % distilled approaches 90%, the intensities of the positively contributing peaks in the LRVs shift to the right to less volatile compounds, suggesting these compounds may contribute more with respect to the predicted ADC temperature. An interesting observation is that some regions (and peaks therein) in the LRVs change sign as the distillation

256 runs toward completion; a good example is a cluster of peaks located ~13 and 17 min in the
257 primary separation dimension and ~1.2 and 1.5 s in the secondary separation dimension.
258 Although the peaks in the LRVs in this separation region are generally positive at 0.025%
259 distilled, as the distillation progresses the magnitude of many peaks diminish until their
260 contribution is zero, then as the distillation progresses further the signs of these peaks change to
261 negative with a corresponding increase in magnitude. This suggests that early in the distillation,
262 analyte compounds corresponding to peaks in the LRV that are changing from positive to
263 negative during the distillation would contribute to increasing the predicted ADC temperature,
264 but approaching the end of the distillation these compounds would contribute to decreasing the
265 predicted ADC temperature. These compounds seem to act analogous to a chemical buffer in
266 that as buffers moderate changes in pH, these compounds moderate the temperature range of the
267 distillation, i.e. the more of these compounds present the narrower the temperature range over
268 which the distillation will occur.

269 An interesting phenomenon is observed at the higher % distilled values, as shown in Fig.
270 3c. There are several unexpected, slightly positive peaks in the LRV region between 5 and 15
271 min. At 90% distilled the chemical composition of the fuels is actually a subset of the fuel
272 composition that is analyzed by the GC × GC – TOFMS instrument, since at 90% distilled the
273 more volatile compounds will have mostly boiled off, and there likely have been some
274 significant changes in the relative compositions of the various compounds in the fuels. Thus, the
275 positive value peaks in the LRVs in the region between 5 and 15 min may be attributed to
276 covariance between compounds that are more volatile and compounds that are less volatile in the
277 PLS models (due to inherent similarities of the RP-1 lab blends), and not necessarily because

these LRV peaks are chemically meaningful; this may lead to a higher source of error in PLS models at higher % distilled values.

Inspection of the negative regions (and peaks therein) in all of the LRVs, analyte compounds between 5 and 15 min generally have negative values, suggesting the earlier and more volatile compounds lower the overall temperature of the ADC at a given % distilled. As with the positive LRV values, as the distillation progresses from 0.025% to 90% distilled, the intensity shifts from left to right. As the temperature rises, the more volatile compounds preferentially evaporate, so their decreased presence reduces their influence on the overall temperature at higher % distilled values, while the heavier, less volatile compounds contribute more. A list of representative, yet key, analyte compounds of interest, indicated by large peak magnitudes in the LRVs were identified and summarized in Tables 2, 3, and 4. For example, methylbutylbenzene (identified in Fig. 1d) is listed in Table 2, and is one of the major positively contributing compounds to the LRV. Decane (identified in Fig. 1b), is listed in Table 3, and is one of the major negatively contributing compounds to the LRV. Adamantane (identified in Fig. 1c) in Table 4 is one of the significant compounds that change sign with respect to their contribution as the ADC nears completion. Identification of compounds that impact the ADC can play an important role in understanding the information provided by the ADC experiment, and ultimately could play a key role in improving fuel formulation and performance.

Finally, we present the LOOCV summary using the RMSECV calculation defined in Eq. (1) as a function of the % distilled value. The LOOCV procedure for the PLS modeling was performed using both sets of GC × GC – TOFMS data with the ADC data set. The most appropriate number of latent variables (LVs) was determined to be 4, based upon the analysis of scree plots [22]. The LOOCV summary in Fig. 4 provides an assessment of the accuracy of the

PLS models. The residuals (Fig. 2d) of many of the RP-1 fuels cross at 80% distilled along with a sharp increase in the RMSECV in Fig. 4 (at 85% and 90% distilled). These changes are linked to the changes in fuel composition as more fuel is distilled and the resulting covariance between compounds of different volatility that appear in the chromatograms. In principle, distillate fractions of RP-1 fuels could be collected at each % distilled and analyzed with the GC × GC – TOFMS, and the resulting chromatograms could be used to construct the PLS models using their respective temperatures on the ADC data. However, this approach is more laborious and impractical, requiring a prohibitively large set of samples, e.g., 190 samples, from 10 fuels x 19 ADC % distilled points (instead of only10 fuel samples directly analyzed herein in order to demonstrate the methodology principles). The primary benefit of collecting and analyzing distillate fractions at each % distilled value would be to reduce the apparent covariance, thus making the RMSECV values (in Fig. 4) consistently smaller across the ADC. Another way to think about this source of the error while approaching the end of the distillation is that PLS is using the chromatograms of un-distilled RP-1 fuels to "predict the future" ADC values. It is likely that better PLS models could be constructed from chromatograms generated from the RP-1 fuels sampled at each % distilled. Using a respective chromatogram of a fuel at each distillation point would have been more representative of the fuel and would have helped minimize the error of the PLS models However, obtaining said RP-1 samples at various stages of distillation poses a significantly more laborious proposition.

## Conclusions

In this report we have demonstrated the use of PLS on GC × GC – TOFMS chromatograms of RP-1 fuels, and their respective ADCs. The PLS modeling provides insight

into how the chemical composition weighs differently in determining the temperature for a given % distilled value across the ADC. Compounds were discovered that correlate with narrowing the temperature range of which the distillation occurs. The predictive power of the PLS modeling assessed using LOOCV was found to be extremely powerful, yielding RMSECV with low values, typically below 2.0 °C, at each % distilled measurement point during the ADC analysis. This outcome bodes well for potential future studies with expanded fuel sample sets.

**Acknowledgement**

The work at the University of Washington (UW) was performed under subcontract to ERC, Incorporated, Air Force Research Laboratory, Edwards AFB, CA. The fuels were provided by the Air Force Research Laboratory/RQRC, Edwards AFB, CA. Certain commercial equipment, instruments or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the University of Washington, the United States Air Force, or the National Institute of Standards and Technology, nor does it imply the materials or equipment identified are necessarily the best available for that purpose.

## References


1. Billingsley MC, Edwards T, Shafer LM, Bruno TJ (2010) Extent and Impacts of Hydrocarbon Fuel Compositional Variability for Aerospace Propulsion Systems, Proc. 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference.

2. Lovestead TM, Windom BC, Riggs JR, Nickell C, Bruno TJ (2010) Assessment of the Compositional Variability of RP-1 and RP-2 with the Advanced Distillation Curve Approach. Energy Fuels 24:5611–5623. doi: 10.1021/ef100994w

3. Gough RV, Bruno TJ (2013) Composition-Explicit Distillation Curves of Alternative Turbine Fuels. Energy Fuels 27:294–302. doi: 10.1021/ef3016848

4. Hsieh PY, Abel KR, Bruno TJ (2013) Analysis of Marine Diesel Fuel with the Advanced Distillation Curve Method. Energy Fuels 27:804–810. doi: 10.1021/ef3020525

5. Burger JL, Bruno TJ (2012) Application of the Advanced Distillation Curve Method to the Variability of Jet Fuels. Energy Fuels 26:3661–3671. doi: 10.1021/ef3006178

6. Begue NJ, Cramer JA, Von Bargen C, Myers KM, Johnson KJ, Morris RE (2011) Automated Method for Determining Hydrocarbon Distributions in Mobility Fuels. Energy Fuels 25:1617–1623. doi: 10.1021/ef101635a

7. Christensen JH, Tomasi G, Hansen AB (2005) Chemical Fingerprinting of Petroleum Biomarkers Using Time Warping and PCA. Environ Sci Technol 39:255–260. doi: 10.1021/es049832d

8. Bruno TJ (2006) Improvements in the Measurement of Distillation Curves. 1. A Composition-Explicit Approach. Ind Eng Chem Res 45:4371–4380. doi: 10.1021/ie051393j

9. Bruno TJ, Smith BL (2006) Improvements in the Measurement of Distillation Curves. 2. Application to Aerospace/Aviation Fuels RP-1 and S-8. Ind Eng Chem Res 45:4381–4388. doi: 10.1021/ie051394b

10. Bruno TJ, Ott LS, Smith BL, Lovestead TM (2010) Complex Fluid Analysis with the Advanced Distillation Curve Approach. Anal Chem 82:777–783. doi: 10.1021/ac902002j

11. Smith BL, Bruno TJ (2007) Improvements in the Measurement of Distillation Curves. 4. Application to the Aviation Turbine Fuel Jet-A. Ind Eng Chem Res 46:310–320. doi: 10.1021/ie060938m

12. Ott LS, Smith BL, Bruno TJ (2008) Advanced distillation curve measurements for corrosive fluids: Application to two crude oils. Fuel 87:3055–3064. doi: 10.1016/j.fuel.2008.04.032

13. Ott LS, Hadler AB, Bruno TJ (2008) Variability of The Rocket Propellants RP-1, RP-2, and TS-5: Application of a Composition- and Enthalpy-Explicit Distillation Curve Method. Ind Eng

Chem Res 47:9225–9233. doi: 10.1021/ie800988u

14. Lovestead TM, Bruno TJ (2009) A Comparison of the Hypersonic Vehicle Fuel JP-7 to the Rocket Propellants RP-1 and RP-2 with the Advanced Distillation Curve Method. Energy Fuels 23:3637–3644. doi: 10.1021/ef900096q

15. Windom BC, Bruno TJ (2011) Assessment of the Composition and Distillation Properties of Thermally Stressed RP-1 and RP-2: Application to Fuel Regenerative Cooling. Energy Fuels 25:5200–5214. doi: 10.1021/ef201077a

16. Smith BL, Bruno TJ (2007) Improvements in the Measurement of Distillation Curves. 3. Application to Gasoline and Gasoline + Methanol Mixtures. Ind Eng Chem Res 46:297–309. doi: 10.1021/ie060937u

17. Windom BC, Bruno TJ (2011) Improvements in the Measurement of Distillation Curves. 5. Reduced Pressure Advanced Distillation Curve Method. Ind Eng Chem Res 50:1115–1126. doi: 10.1021/ie101784g

18. Windom BC, Bruno TJ (2013) Application of Pressure-Controlled Advanced Distillation Curve Analysis: Virgin and Waste Oils. Ind Eng Chem Res 52:327–337. doi: 10.1021/ie302399v

19. Bruno TJ, Wolk A, Naydich A (2009) Stabilization of Biodiesel Fuel at Elevated Temperature with Hydrogen Donors: Evaluation with the Advanced Distillation Curve Method. Energy Fuels 23:1015–1023. doi: 10.1021/ef800740d

20. Huber ML, Lemmon EW, Bruno TJ (2009) Effect of RP-1 Compositional Variability on Thermophysical Properties. Energy Fuels 23:5550–5555. doi: 10.1021/ef900597q

21. Omais B, Courtiade M, Charon N, Thiébaut D, Quignard A, Hennion MC (2011) Investigating comprehensive two-dimensional gas chromatography conditions to optimize the separation of oxygenated compounds in a direct coal liquefaction middle distillate. J Chromatogr A 1218:3233–3240. doi: 10.1016/j.chroma.2010.12.049

22. Kehimkar B, Hoggard JC, Marney LC, Billingsley MC, Fraga CG, Bruno TJ, Synovec RE (2014) Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis. J Chromatogr A 1327:132–140 doi: 10.1016/j.chroma.2013.12.060

23. Hope JL, Sinha AE, Prazen BJ, Synovec RE (2005) Evaluation of the DotMap algorithm for locating analytes of interest based on mass spectral similarity in data collected using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry. J Chromatogr A 1086:185–192. doi: 10.1016/j.chroma.2005.06.026

24. Mohler RE, Dombek KM, Hoggard JC, Young ET, Synovec RE (2006) Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of

Metabolites in Fermenting and Respiring Yeast Cells. Anal Chem 78:2700–2709. doi: 10.1021/ac052106o

25. Pierce KM, Hoggard JC, Mohler RE, Synovec RE (2008) Recent advancements in comprehensive two-dimensional separations with chemometrics. J Chromatogr A 1184:341–352. doi: 10.1016/j.chroma.2007.07.059

26. Hoggard JC, Synovec RE (2008) Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis. Anal Chem 80:6677–6688. doi: 10.1021/ac800624e

27. Fraga CG, Prazen BJ, Synovec RE (2000) Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the High-Speed Quantitative Analysis of Aromatic Isomers in a Jet Fuel Using the Standard Addition Method and an Objective Retention Time Alignment Algorithm. Anal Chem 72:4154–4162. doi: 10.1021/ac000303b

28. Prazen BJ, Johnson KJ, Weber A, Synovec RE (2001) Two-Dimensional Gas Chromatography and Trilinear Partial Least Squares for the Quantitative Analysis of Aromatic and Naphthene Content in Naphtha. Anal Chem 73:5677–5682. doi: 10.1021/ac010637g

29. Johnson KJ, Synovec RE (2002) Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis. Chemom Intell Lab Syst 60:225–237. doi: 10.1016/S0169-7439(01)00198-8

30. Johnson KJ, Prazen BJ, Young DC, Synovec RE (2004) Quantification of naphthalenes in jet fuel with GC×GC/Tri-PLS and windowed rank minimization retention time alignment. J Sep Sci 27:410–416. doi: 10.1002/jssc.200301640

31. Pierce KM, Kehimkar B, Marney LC, Hoggard JC, Synovec RE (2012) Review of chemometric analysis techniques for comprehensive two dimensional separations data. J Chromatogr A 1255:3–11. doi: 10.1016/j.chroma.2012.05.050

32. Hoggard JC, Siegler WC, Synovec RE (2009) Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC. J Chemom 23:421–431. doi: 10.1002/cem.1239

33. Westad F, Afseth NK, Bro R (2007) Finding relevant spectral regions between spectroscopic techniques by use of cross model validation and partial least squares regression. Anal Chim Acta 595:323–327. doi: 10.1016/j.aca.2007.02.015

34. Rajalahti T, Kvalheim OM (2011) Multivariate data analysis in pharmaceutics: A tutorial review. Int J Pharm 417:280–290. doi: 10.1016/j.ijpharm.2011.02.019

35. Gowen AA, Downey G, Esquerre C, O'Donnell CP (2011) Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. J Chemom 25:375–381. doi: 10.1002/cem.1349

36. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:1–17. doi: 10.1016/0003-2670(86)80028-9

37. Mohler RE, Dombek KM, Hoggard JC, Pierce KM, Young ET, Synovec RE (2007) Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software. Analyst 132:756–767. doi: 10.1039/B700061H

38. Pierce KM, Hoggard JC (2013) Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography. Anal Methods. doi: 10.1039/C3AY40965A

39. Marney LC, Christopher Siegler W, Parsons BA, Hoggard JC, Wright BW, Synovec RE (2013) Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data. Talanta 115:887–895. doi: 10.1016/j.talanta.2013.06.038

40. Christian GD (2004) Analytical Chemistry, 6th ed. John Wiley & Sons, New York

## Tables

Table 1. RP-1 Fuel Set [22].

| Sample number | NIST Number [1-2] | AFRL Designation |
|---|---|---|
| 1 | 11 | LB080409-01 |
| 2 | 10 | LB073009-06 |
| 3 | 9 | LB073009-08 |
| 4 | 8 | LB080409-05 |
| 5 | 7 | LB073009-05 |
| 6 | 5 | LB073009-01 |
| 7 | 4 | LB073009-09 |
| 8 | 1 | LB073009-02 |
| 9 | 2 | LB073009-03 |
| 10 | 3 | XC2521HW10 |

Table 2. Major contributing compounds identified in the LRVs that contribute positively, per the blue features in Fig. 3a, b, c. The retention time on the primary column is labeled $^{1}t_R$, and on the secondary column as $^{2}t_R$. The mass spectral match value is labeled MV.

| # | Compound Identification | $^{1}t_R$ (min) | $^{2}t_R$ (s) | MV | Compound Class |
|---|---|---|---|---|---|
| 1 | Trimethyldodecane (C15H32) | 17.42 | 2.23 | 924 | alkanes |
| 2 | 3-Methyltridecane (C14H30) | 19.88 | 2.00 | 910 | alkanes |
| 3 | 3-Methyltetradecane (C15H32) | 20.17 | 1.96 | 922 | alkanes |
| 4 | Heptylcyclohexane (C13H26) | 18.75 | 1.62 | 889 | cycloalkanes |
| 5 | Octylcyclohexane (C14H28) | 21.25 | 1.60 | 909 | cycloalkanes |
| 6 | Nonylcyclohexane (C15H30) | 23.63 | 1.60 | 929 | cycloalkanes |
| 7 | Methyl-bicyclohexyl (C13H24) | 20.79 | 1.36 | 841 | di- & tri- cycloalkanes |
| 8 | Hexamethyloctahydro-1H-indene (C15H28) | 22.21 | 1.43 | 832 | di- & tri- cycloalkanes |
| 9 | Bicyclohexane (C15H28) | 20.00 | 1.38 | 907 | di- & tri- cycloalkanes |
| 10 | Methylbutylbenzene (C11H16) | 19.29 | 0.95 | 908 | mono-aromatics |
| 11 | Azulene (C10H8) | 26.83 | 0.76 | 919 | di-aromatics |

Table 3. Major contributing compounds identified in the LRVs that contribute negatively per the red features in Fig. 3a, b, c.

| # | Compound Identification | $^{1}t_R$ (min) | $^{2}t_R$(s) | MV | Compound Class |
|---|---|---|---|---|---|
| 1 | Methylnonane (C10H22) | 7.46 | 1.94 | 937 | alkanes |
| 2 | Decane (C10H22) | 8.21 | 1.94 | 960 | alkanes |
| 3 | Dimethylnonane (C11H24) | 8.42 | 2.18 | 931 | alkanes |
| 4 | Trimethylcyclohexane (C9H18) | 7.42 | 1.40 | 943 | cycloalkanes |
| 5 | Methylpropylcyclohexane (C10H20) | 8.92 | 1.63 | 873 | cycloalkanes |
| 6 | Ethyldimethylcyclohexane (C10H20) | 9.21 | 1.53 | 864 | cycloalkanes |
| 7 | cis-Octahydro-1H-indene (C9H16) | 11.17 | 1.24 | 948 | di- & tri- cycloalkanes |
| 8 | Dimethylbicyclo[3.2.1]octane (C10H18) | 11.96 | 1.34 | 890 | di- & tri- cycloalkanes |
| 9 | Not found at significant level | | | | mono-aromatics |
| 10 | Not found at significant level | | | | di-aromatics |

Table 4. Compounds of interest identified in the LRVs that exhibit a sign change across the ADC (from positive to negative), per Fig. 3.

| # | Compound Identication | $^{1}t_R$ (min) | $^{2}t_R$ (s) | MV | Compound Class |
|---|---|---|---|---|---|
| 1 | Trimethyldecane (C13H28) | 14.83 | 2.24 | 898 | alkanes |
| 2 | Methyldodecane isomer (C13H28) | 14.67 | 2.10 | 926 | alkanes |
| 3 | Methyldodecane isomer (C13H28) | 15.00 | 2.10 | 940 | alkanes |
| 4 | Not found at significant level | | | | cycloalkanes |
| 5 | Not found at significant level | | | | cycloalkanes |
| 6 | trans-decahydronaphthalene(C10H18) | 12.83 | 1.35 | 930 | di- & tri- cycloalkanes |
| 7 | Adamantane (C10H16) | 15.00 | 1.17 | 959 | di- & tri- cycloalkanes |
| 8 | Methyldecahydronaphthalene (C11H20) | 14.00 | 1.42 | 940 | di- & tri- cycloalkanes |
| 9 | Not found at significant level | | | | mono-aromatics |
| 10 | Not found at significant level | | | | di-aromatics |

## Figure Captions

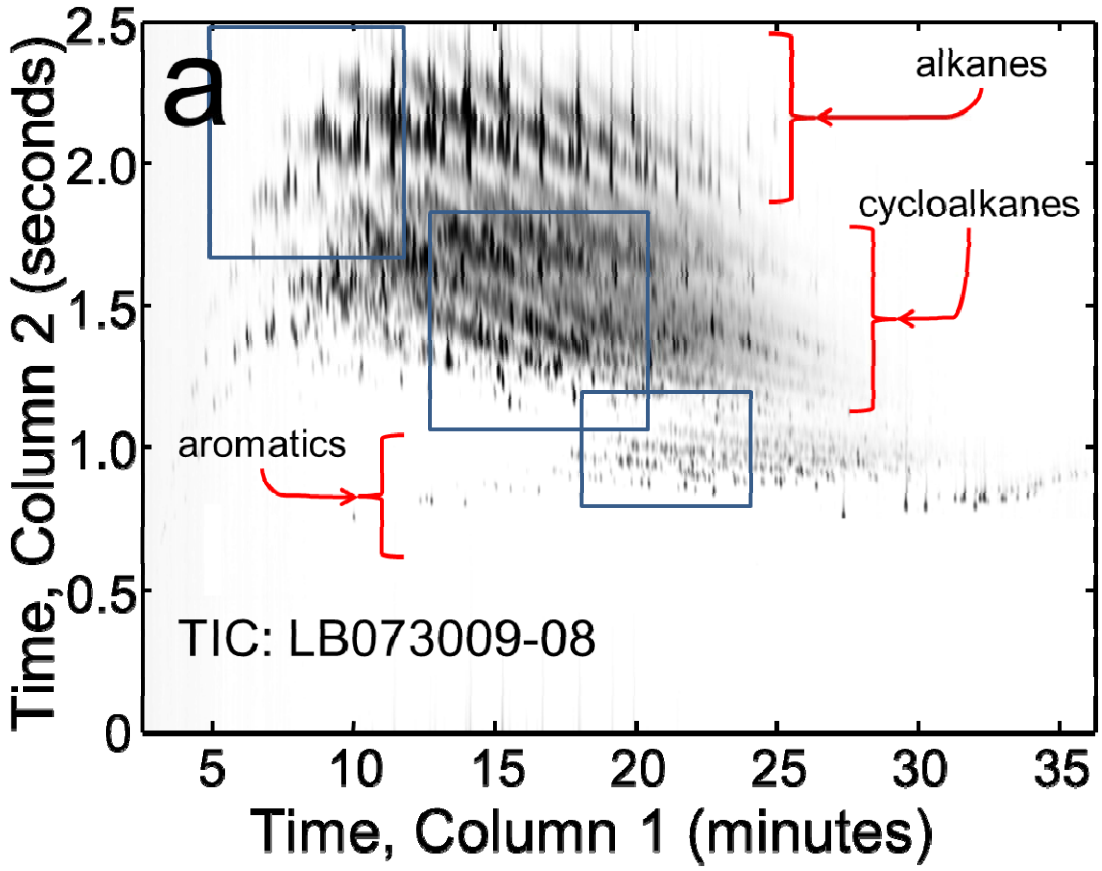
**Fig. 1** (a) Total ion current (TIC) chromatogram of the RP-1 fuel LB073009-08, collected using GC × GC – TOFMS. Compound classes are indicated. (b) Region between 5 min and 12 min in the primary dimension and 1.7 s and 2.5 s in the secondary dimension at m/z 57, the upper left box in (a), with n-decane identified. (c) Region between 13 min and 19 min in the primary dimension and 1.0 s and 1.8 s in the secondary dimension at m/z 136, the middle box in (a), with adamantane identified. (d) Region between 18 min and 24 min in the primary dimension and 0.8 s and 1.2 s in the secondary dimension at m/z 105, the lower right box in (a), with methylbutylbenzene identified.

**Fig. 2** (a) Measured ADC data for the ten RP-1 fuels (listed in Table 1) are provided. (b) The ADC of two RP-1 fuels are provided that span the approximate range of the ADC data set: top LB073009-06, bottom LB073009-02. (c) The PLS modeled ADC for the two fuels in part (b) are provided: top LB073009-06, bottom LB073009-02. (d) The ADC residuals for all ten of the RP-1 fuels, calculated as the predicted ADC obtained from the cross validation predicted PLS models minus the measured ADC.

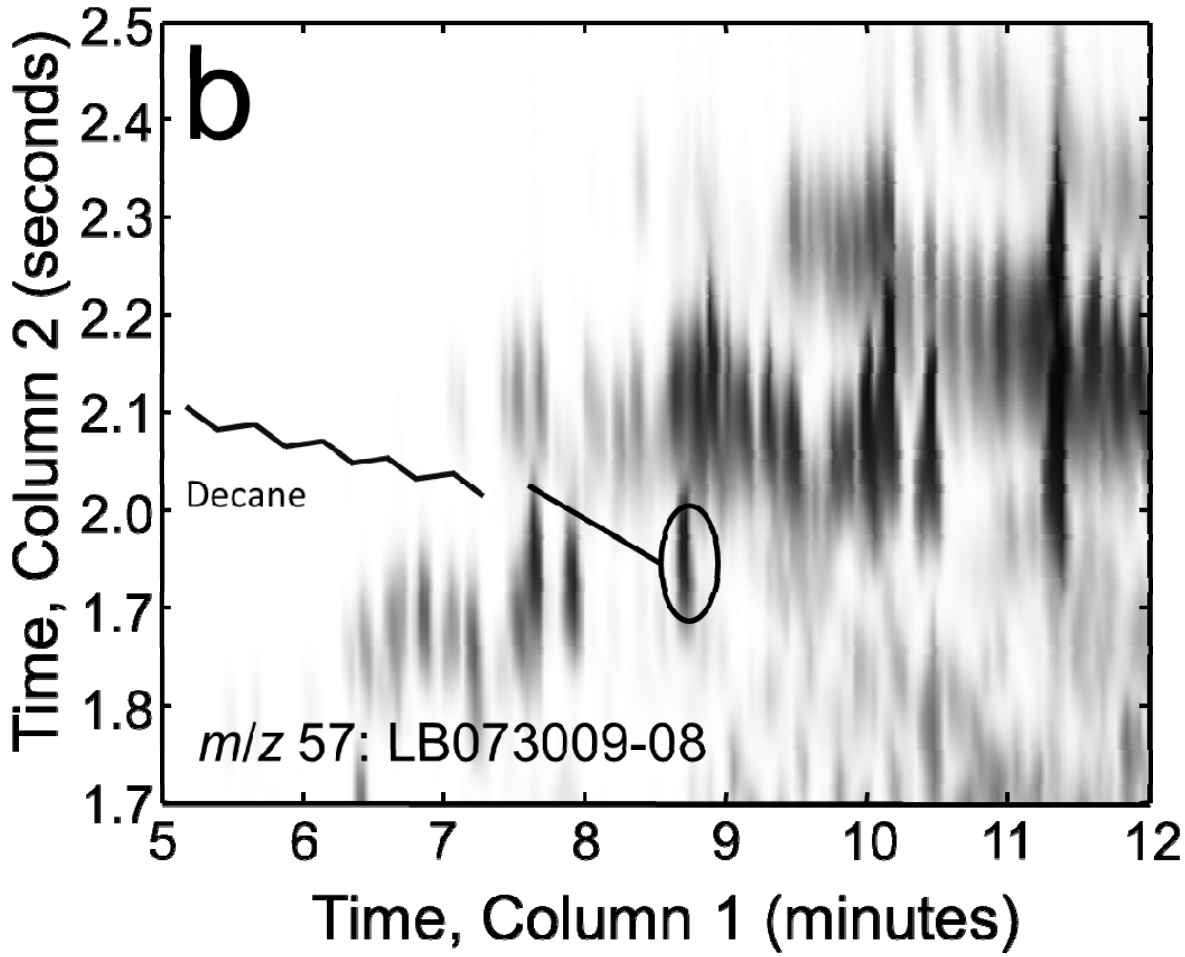
**Fig. 3** (a) Linear regression vector (LRV) of a 4LV PLS model at 0.025% distilled of the ADC, with blue indicating a positive contribution to the LRV and red indicating a negative contribution. (b) LRV of a 4LV PLS model at 45% distilled (the middle) of the ADC. (c) LRV of a 4LV PLS model at 90% distilled (the end) of the ADC.

**Fig. 4** Validation results are provided for the PLS models of the ADCs for the ten RP-1 fuels in Table 1 using LOOCV. The RMSECV values for PLS modeling of both sets of GC × GC – TOFMS data are indicated as a function of % distilled.

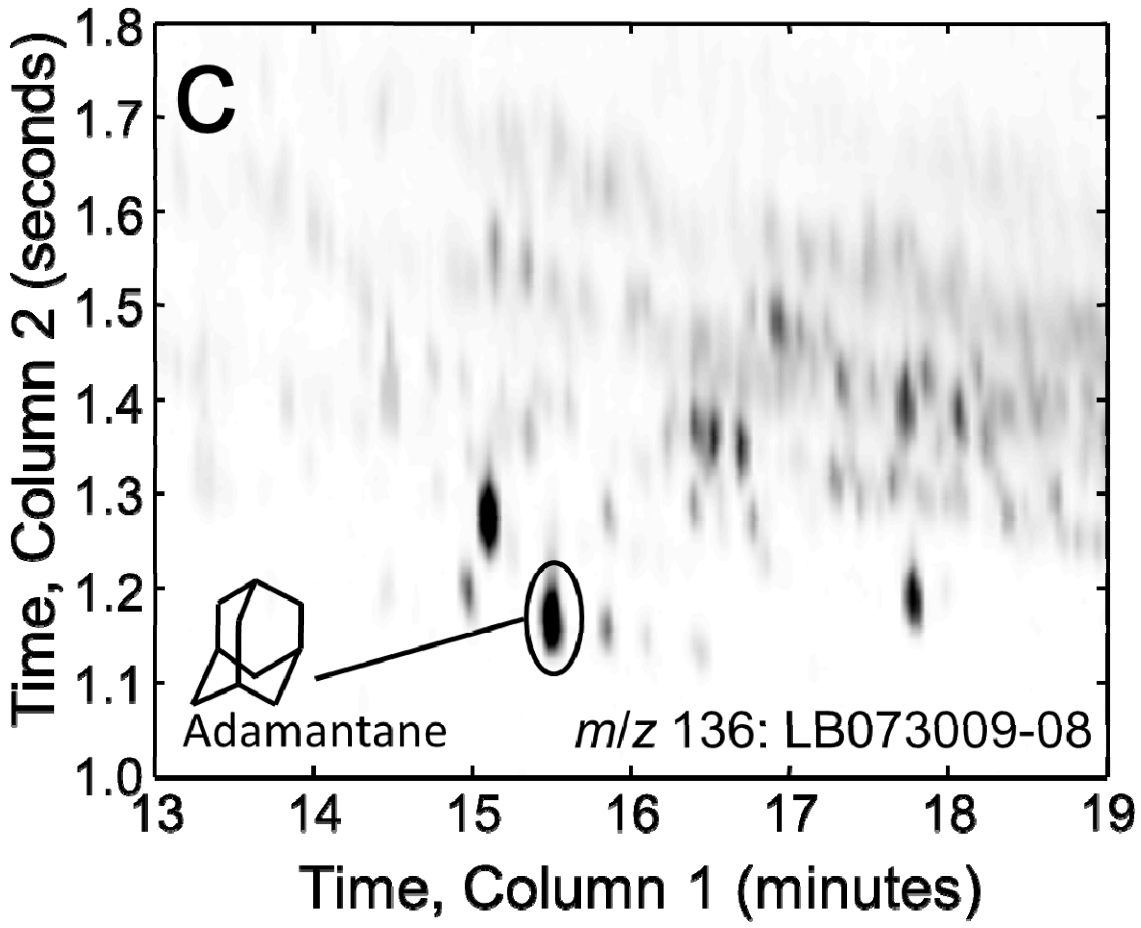
# Figure 1A



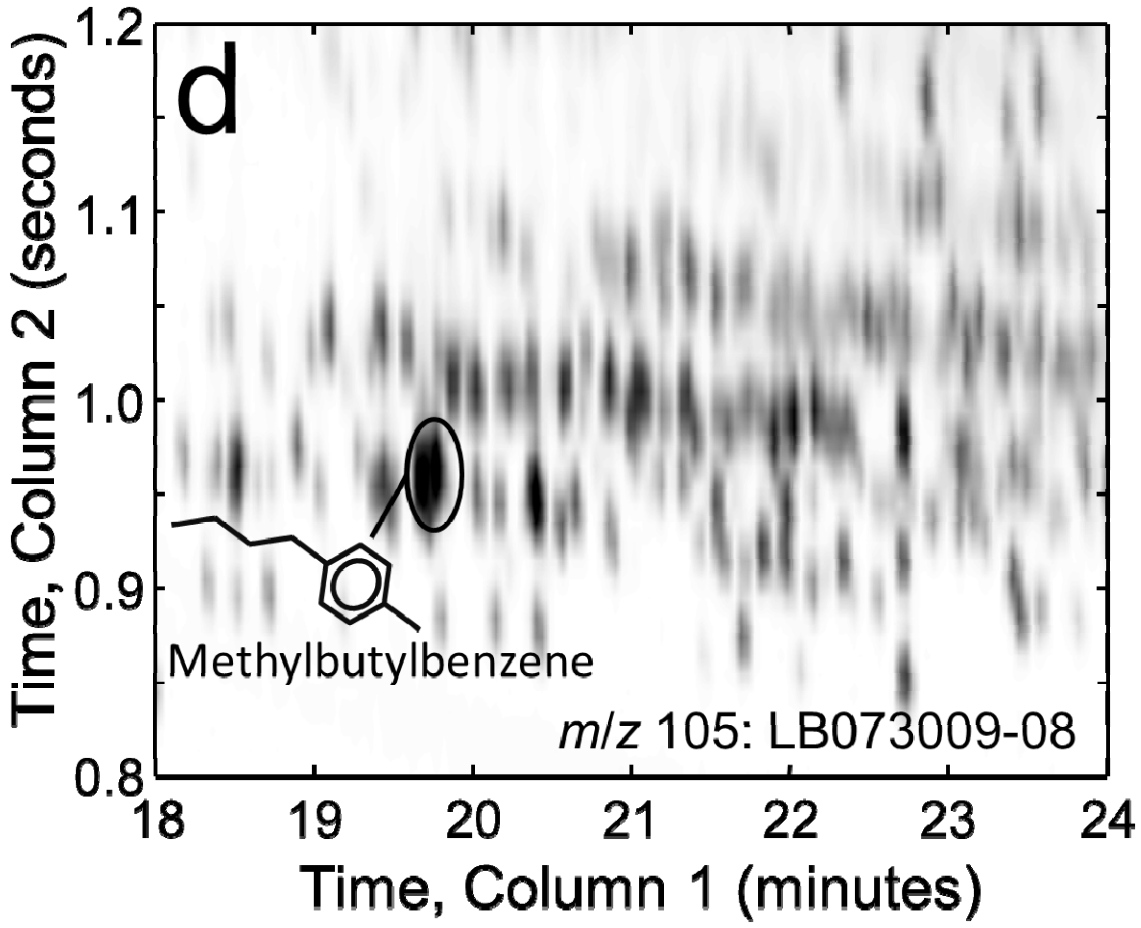
# Figure 1B



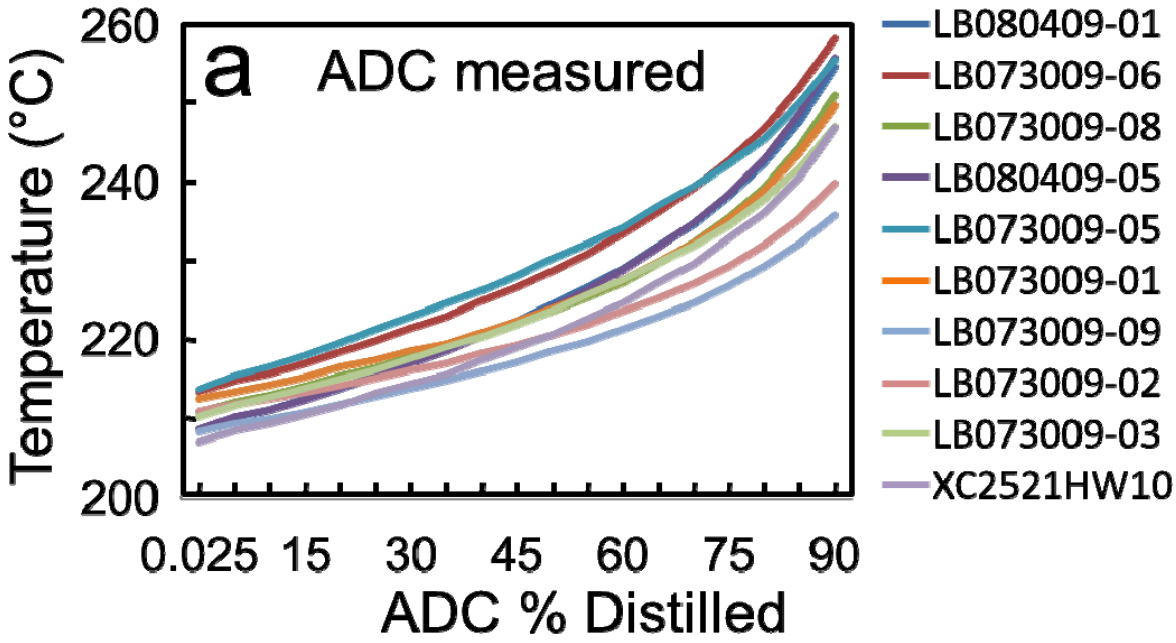
# Figure 1C



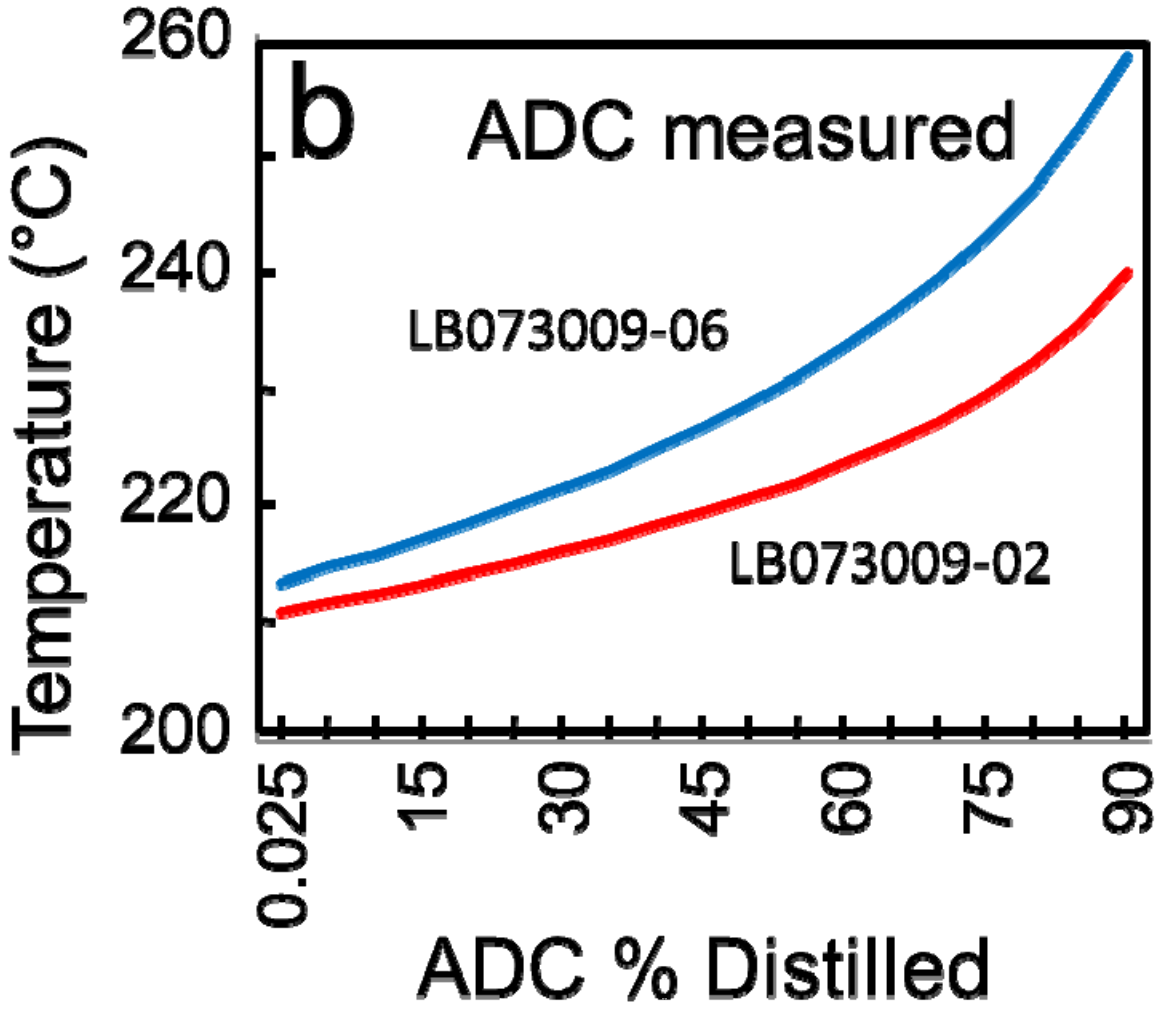
# Figure 1D



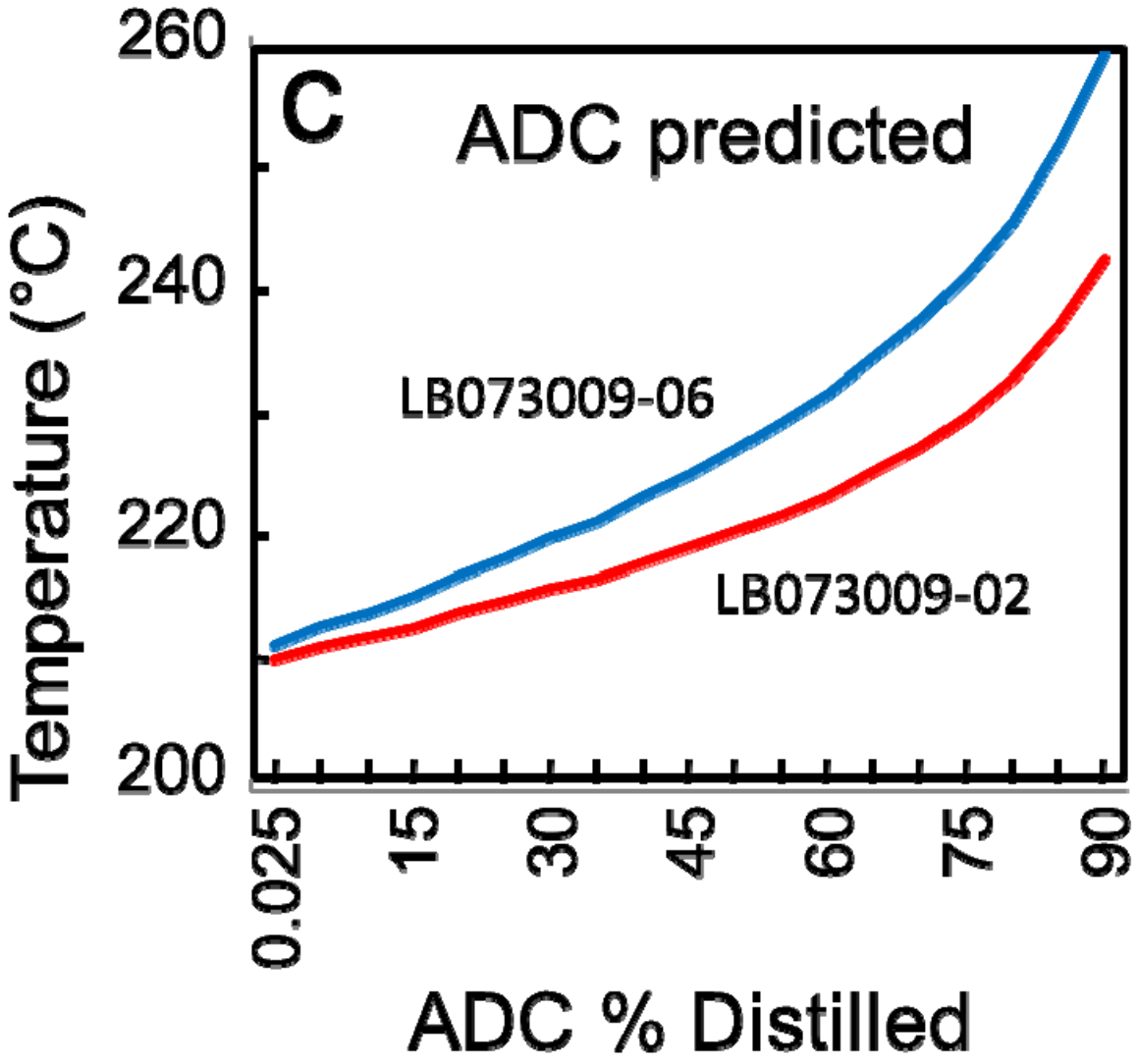
# Figure 2A



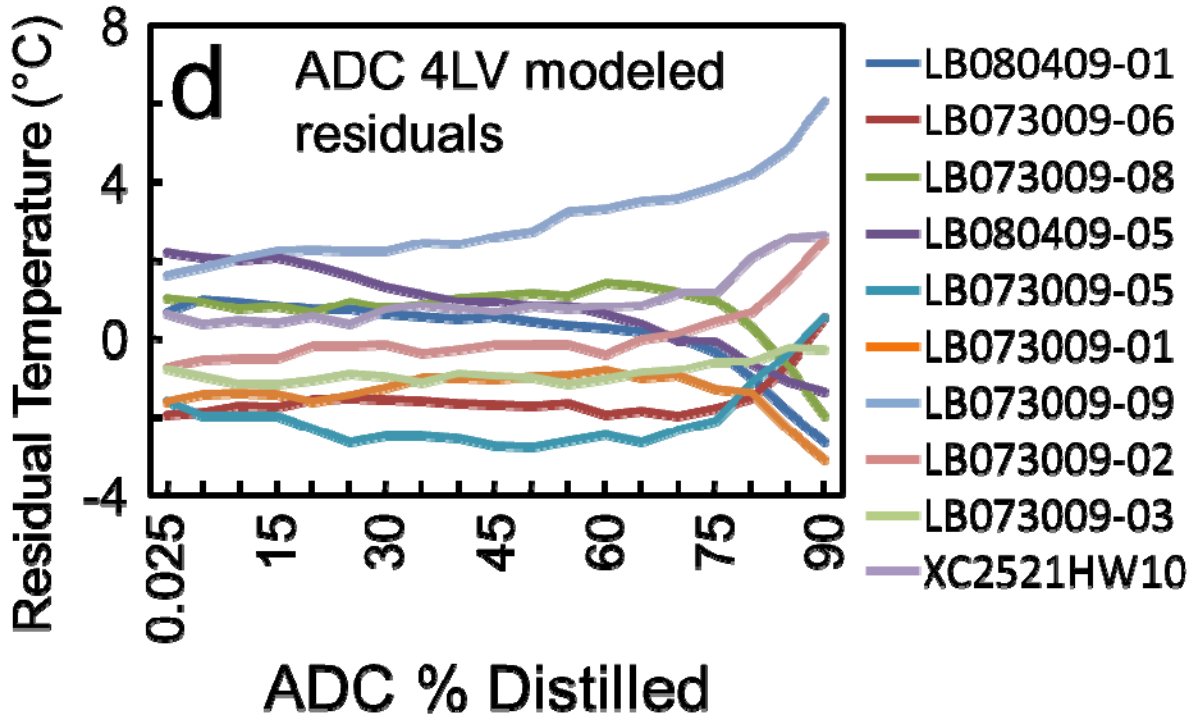
# Figure 2B



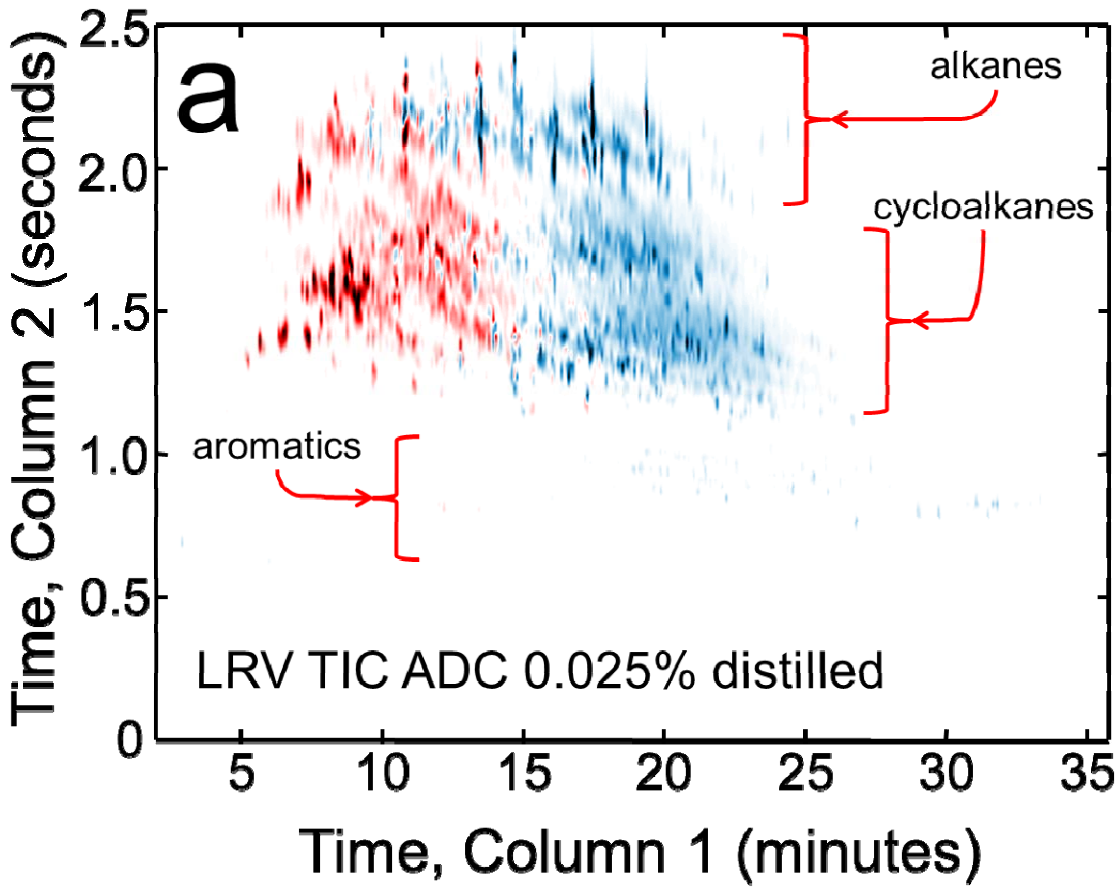
# Figure 2C



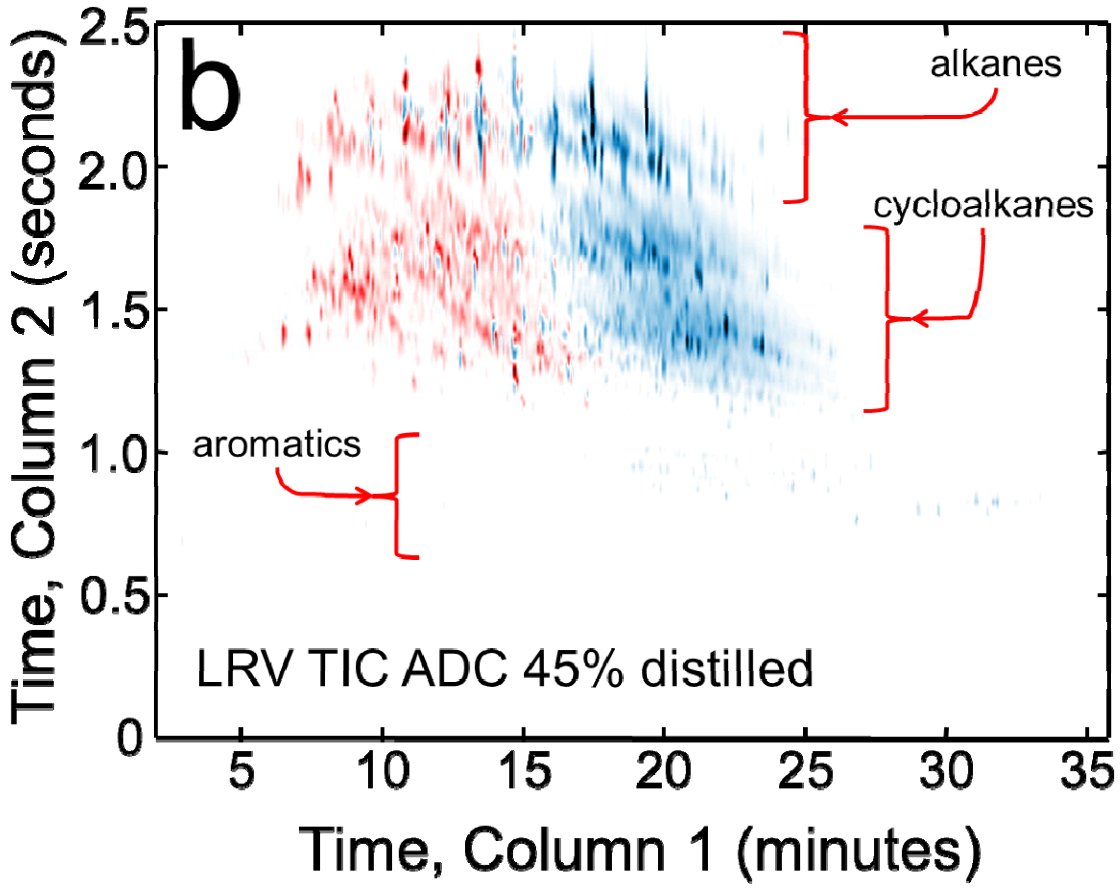
# Figure 2D



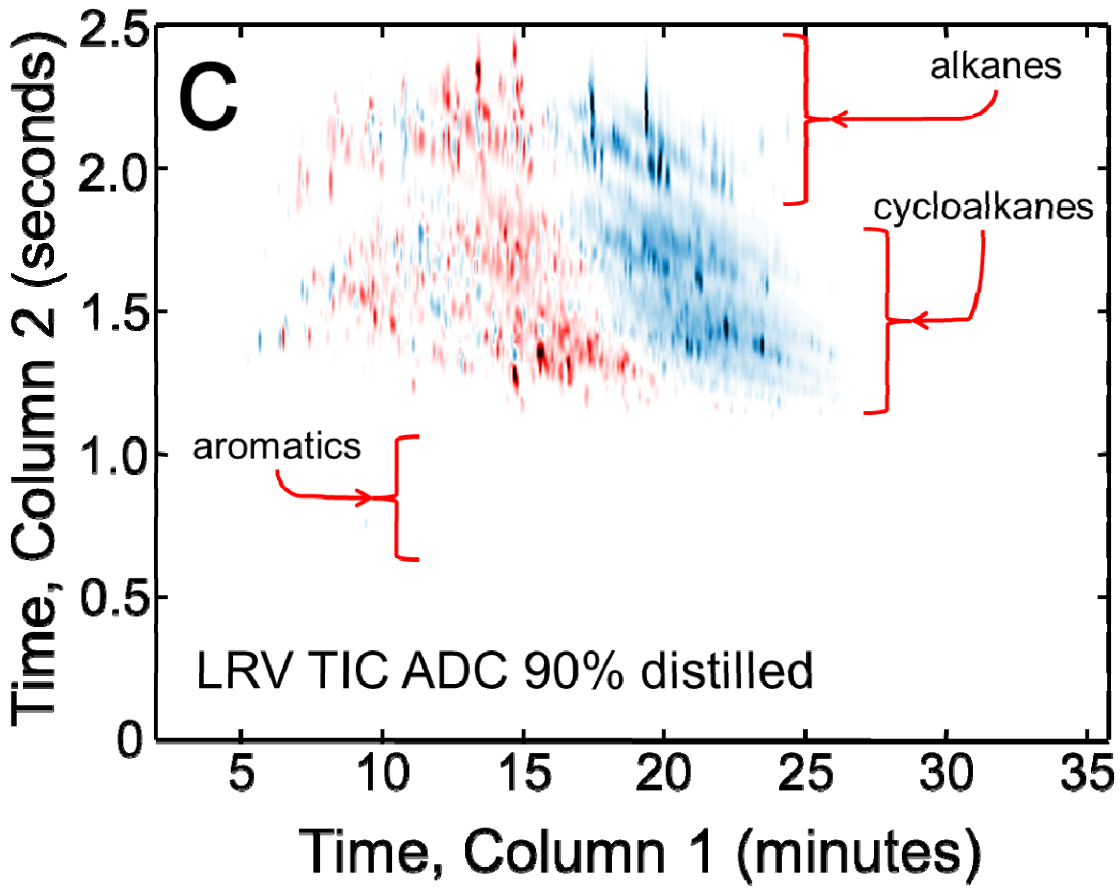
# Figure 3A



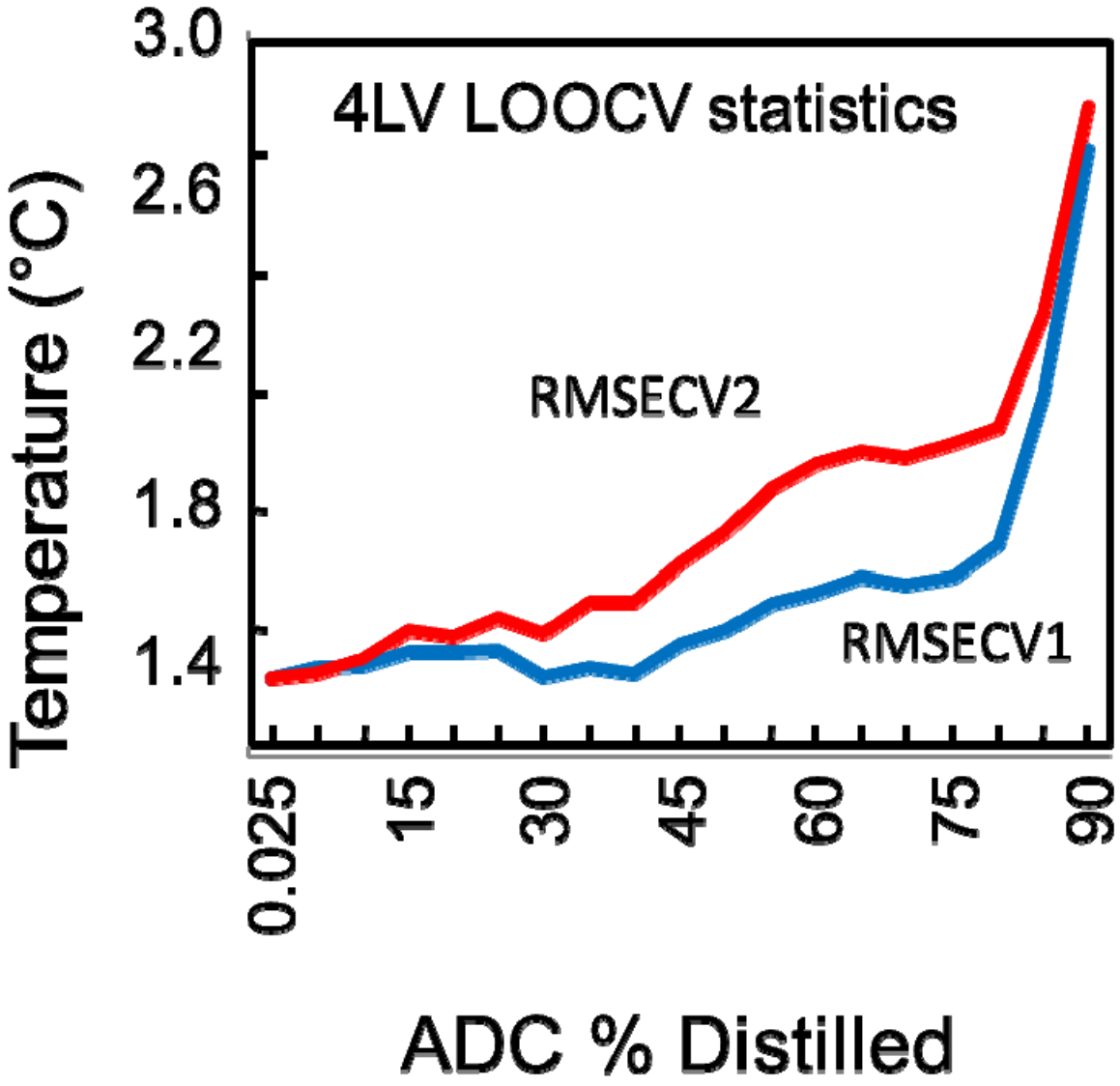
# Figure 3B



# Figure 3C



780 # Figure 4




781